Cancer Causes and Control **15**: 921–929, 2004.

Geographic clustering of residence in early life and subsequent risk of breast cancer (United States)

Daikwon Han[1,2,*], Peter A. Rogerson[2,3], Jing Nie[1], Matthew R. Bonner[1], John E. Vena[4], Dominica Vito[1], Paola Muti[1], Maurizio Trevisan[1], Stephen B. Edge[5] & Jo L. Freudenheim[1]
[1]*Department of Social and Preventive Medicine, University at Buffalo, Buffalo, NY 14214, USA;* [2]*Department of Geography and National Center for Geographic Information and Analysis, University at Buffalo, Buffalo, NY 14261, USA;* [3]*Department of Biostatistics, University at Buffalo, Buffalo, NY 14214, USA;* [4]*Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC 29208, USA;* [5]*Department of Breast and Soft Tissue Surgery, Roswell Park Cancer Institute, Buffalo, NY 14263, USA*

Received 23 March 2004; accepted in revised form 3 June 2004

Key words: breast cancer, early-life exposure, spatial clustering.

Abstract

Objective: This study focused on geographic clustering of breast cancer based on residence in early life and identified spatio-temporal clustering of cases and controls.
Methods: Data were drawn from the WEB study (Western New York Exposures and Breast Cancer Study), a population-based case–control study of incident, pathologically confirmed breast cancer (1996–2001) in Erie and Niagara counties. Controls were frequency-matched to cases on age, race, and county of residence. All cases and controls used in the study provided lifetime residential histories. The k-function difference between cases and controls was used to identify spatial clustering patterns of residence in early life.
Results: We found that the evidence for clustered residences at birth and at menarche was stronger than that for first birth or other time periods in adult life. Residences for pre-menopausal cases were more clustered than for controls at the time of birth and menarche. We also identified the size and geographic location of birth and menarche clusters in the study area, and found increased breast cancer risk for pre-menopausal women whose residence was within the cluster compared to those living elsewhere at the time of birth.
Conclusion: This study provides evidence that early environmental exposures may be related to breast cancer risk, especially for pre-menopausal women.

Introduction

Breast cancer is one of the leading causes of death among women in the United States. However, the epidemiology of breast cancer is not yet fully understood. We also do not fully understand mechanisms for the known risk factors; for instance, why changes in age at menarche or age at first birth have an impact on breast cancer risk. A substantial degree of geographical variation in breast cancer incidence and mortality in the US has been observed [1, 2]. While inconclusive, several environmental risk factors are also believed to be involved in breast cancer incidence [3, 4]. There is speculation that environmental factors may explain geographic variation in breast cancer rates not explained by known risk factors. For this reason, the potential role of environmental exposures in breast cancer risk is of particular interest.

In addition, there is a growing interest in early life and lifetime exposures in relation to breast cancer risk. The life course approach is of interest because there may be sensitive time periods for exposures and/or there may be cumulative effects of lifetime exposure involved in breast cancer incidence [5, 6]. Early life has an effect on breast cancer etiology evidenced by the known risk factors

*Address correspondence to: Daikwon Han, Department of Social and Preventive Medicine, University at Buffalo, 3435 Main St., Farber Hall, Rm 270, Buffalo, NY 14214, USA. Ph.: +1-716-829-2975 ext. 605; Fax: +1-716-829-2979; e-mail: dhan@buffalo.edu

such as age at menarche, age at first birth and parity. There is new evidence that even earlier exposures may have an impact on adult breast cancer risk [7]. Trichopoulos [8] suggested that the *in-utero* and perinatal period might be pathologically significant and that the risk of adult breast cancer could be related to high estrogen exposure in early life. There is also accumulating evidence that factors related to early exposure, such as birthweight, may be related to risk [9, 10].

There has been little research investigating possible effects of environmental exposures in early life on subsequent breast cancer risk. Using residence as a proxy measure for environmental exposures, we investigated whether there was any evidence of geographic clustering of adult breast cancer cases associated with their residences in early life. Clustering analyses have often been used to provide clues for the unknown etiology of disease, and thus to generate hypotheses for further epidemiologic research [11]. We looked at the geographic clustering of residence at early critical time points: at birth, at menarche, and at the woman's first birth. By comparing differences in clustering patterns between case and control residences, we were interested in identifying time periods critical to potential environmental exposures and subsequent breast cancer risk.

Methods

Population-based case–control study of breast cancer

We conducted a case–control study of breast cancer in western New York – the WEB study (Western New York Exposures and Breast Cancer Study) . Cases were women, age 35–79 with incident, primary, pathologically confirmed breast cancer diagnosed in Erie and Niagara counties during the period 1996–2001, with no previous cancer diagnosis other than non-melanoma skin cancer. Controls were frequency matched to cases on age, race, and county of current residence; controls under 65 years of age were randomly selected from a New York State Department of Motor Vehicles list and those 65 years and over were chosen from a Health Care Finance Administration list. We ascertained cases by having a nurse–case finder visit the pathology departments of almost all hospitals in these counties. One hospital which did not participate does almost no cancer surgery and refers patients to other participating hospitals. For the one other hospital that did not participate, breast cancer cases were identified in the practice of the breast surgeons who see more than 99% of the cases from that hospital. Extensive in-person interviews and self-administered questionnaires were used to ascertain lifetime residential history and other breast cancer risk factors. A total of 1166 cases and 2105 controls were interviewed. Response rates were 72 and 65% for cases and controls, respectively.

All participants were asked to complete a lifetime residential history, to list the street address, town/city and zip code for their current address and then all other previous addresses throughout their lifetime. Participants provided 20,240 addresses, an average of approximately six addresses for each individual. In this study we focused on residence at the time of the participants' birth, menarche, and at the time that she had her first birth. Analyses were restricted to women residing in Erie or Niagara counties at each of these time points. There were, of course, participants whose addresses were the same for two or more of these times.

For women with incomplete residential information, additional information was obtained using historical city directories. We used these directories to find old addresses, and utilized various resources, such as web searches and commercial address databases for recent addresses. We also examined validity and reliability of reports of earlier residences in a number of ways. For birth addresses, we asked for information on birth address twice and have collected information on reliability of response. For the other time periods, we used information on maiden name and partial address information provided by the participants to search for records in city directories for the appropriate time periods. To improve our ability to geocode addresses, we developed several strategies. First, all addresses were standardized to be matched with the standard format used in GIS. We used the enhanced version of TIGER (Topologically Integrated Geographic Encoding and Referencing Systems), GDT/Dynamap 2000 [12], and overall matching rates were improved about 15–20 % when compared with the use of TIGER as a reference theme. We also used the stand-alone address cleaner ZP4 (Semaphore Co.) to correct and update zip code information to be matched with United States Postal Services certified addresses.

More than 85% of addresses were geocoded using the above strategies and resources. We failed to geocode some addresses primarily because of missing residential information, such as missing street numbers or street names. Since we are dealing with historical residential information, the likelihood of missing previous residential information was higher than that for current residential information. Table 1 is a summary table showing the numbers of cases and controls with complete residential information who resided in the two counties for each of the time periods. The percentage of missing residential information associated with

Table 1. Residential history of breast cancer cases and controls: numbers and percentage of complete and missing residences in Erie and Niagara counties: WEB Study, 1996–2001

	Complete residence		Incomplete or missing residence		Total eligible Erie and Niagara county residence at each time period	
	Case	Control	Case	Control	Case	Control
Birth	505 (79.9%)	804 (81.0%)	127 (20.1%)	189 (19.0%)	632	993
Menarche	673 (87.3%)	1143 (88.1%)	98 (12.7%)	154 (11.9%)	771	1297
First birth	616 (86.4%)	1153 (87.3%)	97 (13.6%)	167 (12.7%)	713	1320

each early life event was highest for birth addresses, at about 20%.

Clustering analyses of residences

To compare clustering patterns of breast cancer cases and controls at each time period, the primary method used was based on the *k*-function [13]. The *k*-function for a point process is defined as the number of events within distance h of an arbitrary event, divided by the overall intensity of events. It is estimated by

$$\lambda\hat{k}(h) = \sum_{i=1}^{n}\sum_{j=1}^{n} w(s_i, s_j)^{-1} I(d_{ij} \leq h)/n, \quad h > 0$$

where n is the number of events, λ is the expected density of events in the study region, h is the pre-specified distance, d_{ij} is the Euclidian distance between point i and point j, I is an indicator function that is equal to one if inter-event distances (d_{ij}) are less than or equal to h, and zero otherwise, and $w(s_i,\ s_j)$ is an edge correction estimator which is the proportion of the circumference of a circle centered at s_i, passing through s_j and that is inside the study area A [14]. Under the null hypothesis of spatial randomness, the expected value of $k(h)$ is πh^2. Geographic clustering will yield values of the *k*-function that are greater than this, since clustering will result in more pairs of points separated by a distance of h than would be expected in a random pattern.

We used the difference between *k*-functions for cases and controls to compare two patterns (*i.e.*, $D(h) = k_{\text{case}}(h) - k_{\text{control}}(h)$). Positive values of $D(h)$ indicate spatial clustering of cases relative to the spatial clustering of controls. Under the null hypothesis of random labeling of cases and controls, the expected value of $D(h)$ is zero, indicating that the *k*-functions of the cases and controls are the same. The test statistic, $D(h)$, was calculated with confidence envelopes using the *splancs* library in *S-plus* [15]. We obtained the approximate 95% confidence limits for two standard errors ($\pm\, 2\sqrt{Var\{D(h)\}}$) at the $\alpha = .05$ level [16]. When the estimated function $D(h)$ deviated from zero by greater than two standard deviations, we interpreted this as a statistically significant difference between the case and control patterns.

We also employed a spatial clustering method to identify significant geographic clusters of breast cancer cases. The spatial scan statistic [17], which considers the likelihood of observing the actual number of cases inside of a circle under the null hypothesis of no clustering, was applied to residence at early life events. We were mainly interested in spatial clustering of high rates, and employed the Bernoulli model based on the locations of individual cases and controls [18]. In addition, odds ratios (OR) and 95% confidence intervals (95% CI) were obtained using logistic regression, adjusting for age, education, age at menarche, parity, history of benign breast disease, family history of breast cancer. All analyses were conducted for the entire group of study participants and for data stratified on menopausal status. Women were considered post-menopausal if their menses had ceased permanently and naturally. Among other women, participants were also considered post-menopausal if any of the following conditions were true: they were on hormone replacement therapy and were over age 55, they had had a bilateral oophorectomy, they had had a hysterectomy without removal of the ovaries and they were older than 50, their menses had ceased permanently due to radiation or other medical treatment and they were older than 55.

Results

Characteristics of subjects included in the analysis, subjects with missing residential information, and subjects excluded due to residence outside of Erie and Niagara counties, are shown in Table 2. About half of the sample was excluded for each time period; the highest percentage of ineligible cases and controls was at the birth residence (46 and 51% respectively). However, we found little difference in characteristics between

Table 2. Characteristics of subjects included in the analysis, subjects with missing residential information, and subjects excluded due to residence outside of the study area (Mean ± SD): WEB Study, 1996–2001

	Cases (n = 1166)			Controls (n = 2105)		
	Included	Missing	Ineligible*	Included	Missing	Ineligible*
Birth	(n = 505)	(n = 127)	(n = 534)	(n = 804)	(n = 189)	(n = 1112)
Age (years)	56.5 ± 10.9	60.0 ± 11.0	58.9 ± 11.3	55.6 ± 11.7	58.0 ± 11.8	59.4 ± 11.7
Education (years)	13.5 ± 2.4	13.1 ± 2.5	13.6 ± 2.7	13.4 ± 2.2	13.2 ± 2.2	13.3 ± 2.5
Parity	2.2 ± 1.5	2.4 ± 1.7	2.4 ± 1.8	2.6 ± 1.8	2.7 ± 1.8	2.8 ± 1.8
Age at menarche (years)	12.4 ± 1.5	12.6 ± 1.5	12.7 ± 1.7	12.7 ± 1.7	12.6 ± 1.6	12.7 ± 1.7
Age at first birth (years)	24.3 ± 4.6	23.5 ± 4.5	24.2 ± 5.1	24.5 ± 4.3	23.5 ± 4.2	24.0 ± 4.7
Pre-menopausal (%)	35.2	18.9	26.4	31.7	28.6	24.6
Body Mass Index	28.2 ± 6.4	28.4 ± 5.8	28.7 ± 6.4	28.0 ± 6.2	28.2 ± 6.0	28.4 ± 6.4
Family history of breast cancer (% yes)	21.3	18.9	20.2	12.7	16.2	12.4
History of benign breast disease (% yes)	34.9	37.0	32.8	22.3	25.9	20.6
Menarche	(n = 673)	(n = 98)	(n = 395)	(n = 1143)	(n = 154)	(n = 808)
Age (years)	56.6 ± 10.7	60.1 ± 11.6	59.5 ± 11.3	56.0 ± 11.7	60.2 ± 11.7	59.9 ± 11.6
Education (years)	13.5 ± 2.4	12.8 ± 2.6	13.6 ± 2.8	13.4 ± 2.2	13.0 ± 2.3	13.3 ± 2.6
Parity	2.2 ± 1.6	2.8 ± 1.8	2.5 ± 1.8	2.6 ± 1.8	2.9 ± 2.1	2.9 ± 1.8
Age at menarche (years)	12.5 ± 1.6	12.8 ± 1.5	12.7 ± 1.7	12.7 ± 1.6	12.6 ± 1.7	12.7 ± 1.7
Age at first birth (years)	24.3 ± 4.6	23.0 ± 4.3	24.2 ± 5.3	24.4 ± 4.5	23.8 ± 4.4	24.0 ± 4.6
Pre-menopausal (%)	30.3	24.5	24.6	33.8	23.4	23.3
Body Mass Index	28.1 ± 6.2	29.5 ± 6.4	28.7 ± 6.5	28.3 ± 6.5	27.6 ± 5.5	28.2 ± 6.1
Family history of breast cancer (% yes)	20.2	22.4	20.6	13.1	13.2	12.1
History of benign breast disease (% yes)	34.5	40.8	31.9	22.3	19.5	21.3
First Birth	(n = 616)	(n = 97)	(n = 453)	(n = 1153)	(n = 167)	(n = 785)
Age (years)	57.4 ± 11.1	58.9 ± 10.8	58.5 ± 11.2	57.0 ± 11.7	60.6 ± 10.7	58.5 ± 12.0
Education (years)	13.4 ± 2.3	13.0 ± 2.9	13.7 ± 2.8	13.3 ± 2.2	13.0 ± 2.1	13.4 ± 2.6
Parity	2.7 ± 1.3	3.1 ± 1.5	1.7 ± 1.9	3.0 ± 1.5	3.4 ± 1.7	2.2 ± 2.0
Age at menarche (years)	12.6 ± 1.5	12.5 ± 1.8	12.6 ± 1.6	12.7 ± 1.6	12.6 ± 1.5	12.7 ± 1.7
Age at first birth (years)	24.8 ± 4.8	22.2 ± 4.1	23.4 ± 4.9	24.7 ± 4.6	22.9 ± 3.5	23.3 ± 4.4
Pre-menopausal (%)	29.4	26.8	26.0	32.2	18.0	26.6
Body Mass Index	28.4 ± 6.3	30.1 ± 6.6	28.2 ± 6.3	28.1 ± 6.1	28.3 ± 6.5	28.4 ± 6.4
Family history of breast cancer (% yes)	21.2	23.7	18.6	11.5	19.4	13.6
History of benign breast disease (% yes)	34.7	37.1	32.7	21.0	25.1	22.0

* Ineligible due to residence outside of Erie and Niagara county.

those subjects included and those subjects with addresses outside of these two counties.

Mapping was used to identify geographic patterns of breast cancer cases and controls for each of the early life events. Maps showing the locations of cases and controls in Figure 1 portray the underlying geographic patterns of breast cancer cases and controls in the study area. The rectangular region was used instead of the actual county boundary as an approximate boundary of the study area to protect individuals' confidentiality. The purpose of such mapping is to inspect patterns visually – the first step in any spatial analysis. Geographic patterns do not appear to vary much from one time period to the next, and they appear to reflect patterns of population distribution in the study area. However, it is difficult to determine whether they were clustered or dispersed relative to population from visual inspection alone, because of the large number of data points.

To assess potential effects of geographic selection bias in our study, we also examined the distribution of current residence in relation to other population data on the geographic distribution of breast cancer cases and the general population. We did not find differences in the geographic distribution of participating and non-participating cases, or between controls and the underlying population, except some tendency for both cases and controls living closer to the interview site to be somewhat more likely to participate than those living further away.

Spatial clustering of residences associated with early life events

We obtained differences between the case and control patterns for locations associated with each early life event. The k-function differences for values of h up to 15 miles, with approximate 95% confidence envelopes, are shown

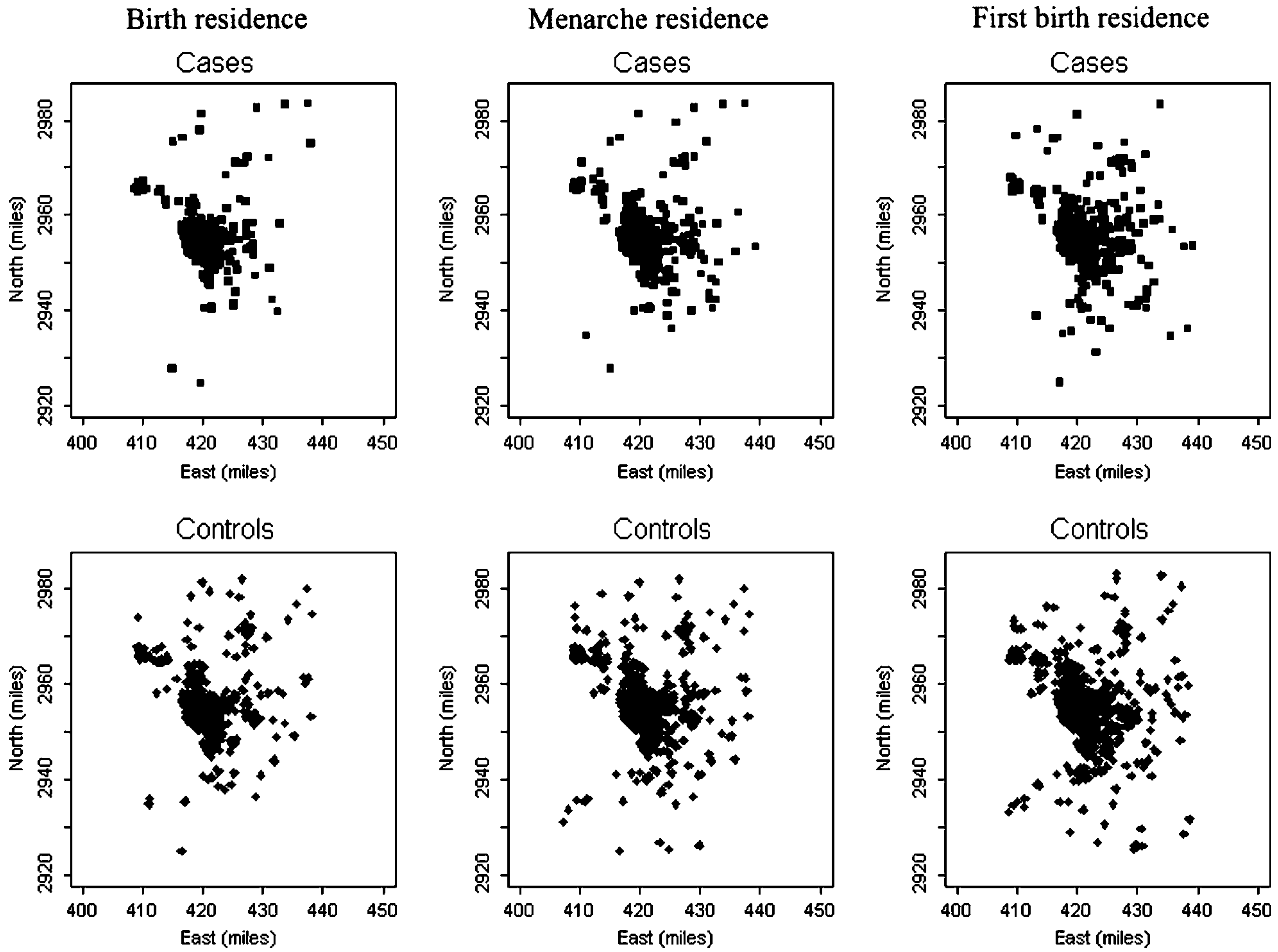

Fig. 1. Residential location of breast cancer cases and controls at each time period: WEB Study, 1996–2001.

in Figure 2. The maximum value of h is generally taken as one-third of the linear extent of the study area [19]. Any patterns beyond this scale can be disregarded, since either peaks or troughs in this geographic scale are difficult to interpret, and are potentially misleading. Figure 2a shows k-function differences for birth residence. It is clear that the estimated function shows strong evidence of spatial clustering, that is, of clustering of cases relative to controls. There was no significant difference up to three miles; statistically significant differences were detected beyond the scale of three miles. There is also evidence of some degree of clustering for breast cancer cases at menarche residence (Figure 2b). Estimates of the D-function are positive but not statistically significant up to seven miles; spatial clustering of breast cancer cases occurs at a scale of about 7–15 miles. For residence at women's first birth and for current residence, the difference is not statistically significant; the plot falls within the confidence interval over all distances (Figures 2c and d).

To determine whether there are any differences in clustering patterns by menopausal status, the k-function difference was performed for pre-menopausal and post-menopausal women separately (Figure 3). We found significant clustering of pre-menopausal breast cancer cases compared to controls for both birth and menarche residence (Figures 3a), while there is no evidence of clustering for post-menopausal breast cancer cases for either period (Figures 3b). We did not find evidence of clustering for first birth and current residence (at diagnosis) for either group (not shown). Estimated functions at birth residence show a strong clustering of pre-menopausal cases over the entire geographic scale with a peak at seven miles. Values are positive for post-menopausal cases, but not statistically significant. For menarche residence, we also observed a strong clustering of pre-menopausal cases with a peak at about 8–10 miles. Again differences are not statistically significant for post-menopausal women at menarche residence.

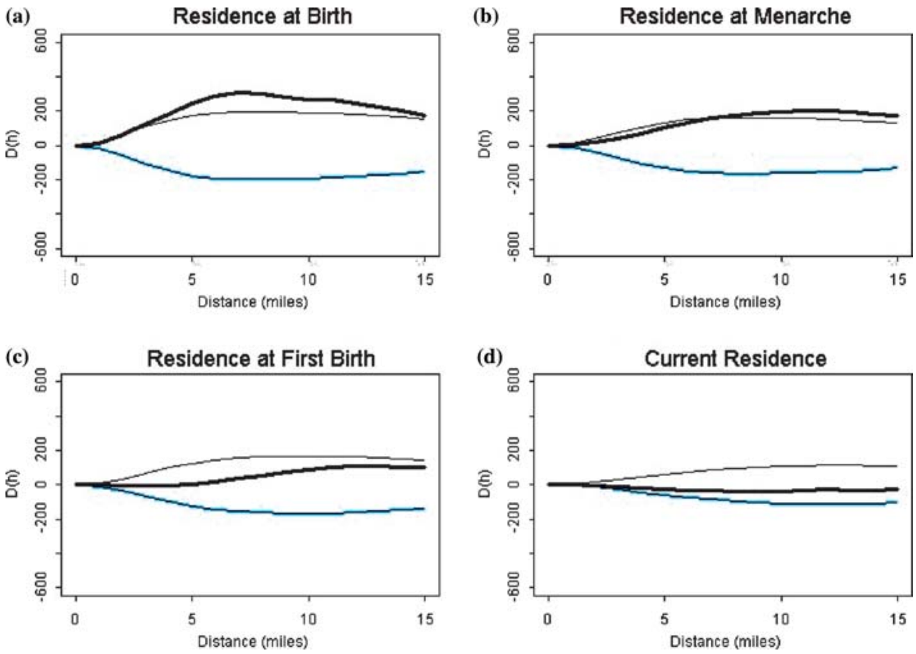

Fig. 2. k-function differences in clustering patterns between breast cancer cases and controls, WEB Study, 1996–2001: shown are *k*-function differences in black and 95% confidence limits in grey.

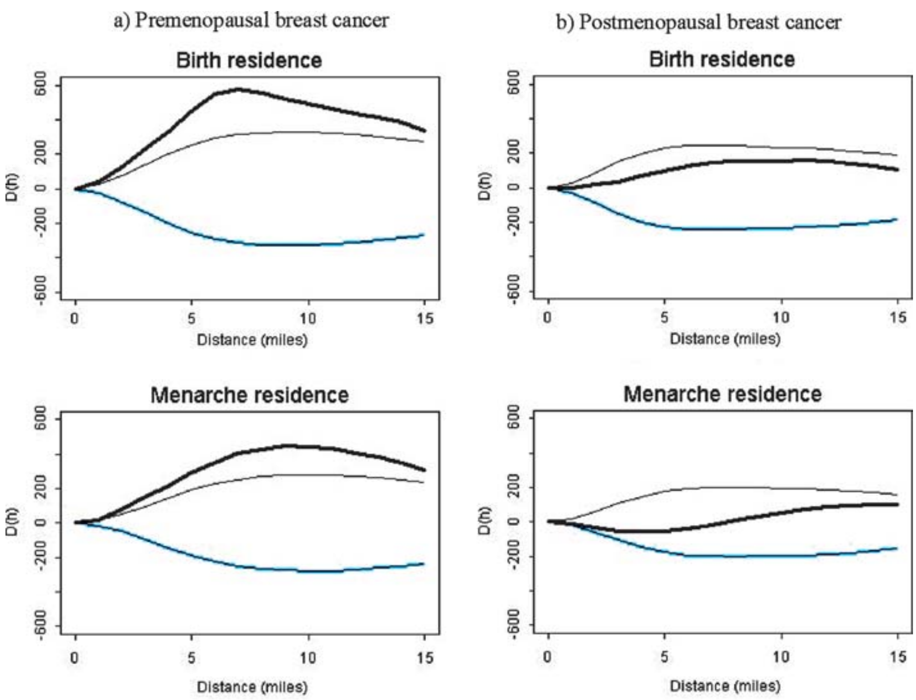

Fig. 3. k–function differences in clustering patterns between breast cancer cases and controls by menopausal status, WEB Study, 1996–2001.

Identifying the geographic location of breast cancer clusters

To identify the geographic location of areas with higher intensities for pre-menopausal cases in the study area, the spatial scan statistic was applied to residences of pre-menopausal women at the time of birth and menarche. Maps in Figure 4 present results of the clustering analysis. The circle in Figure 4a indicates

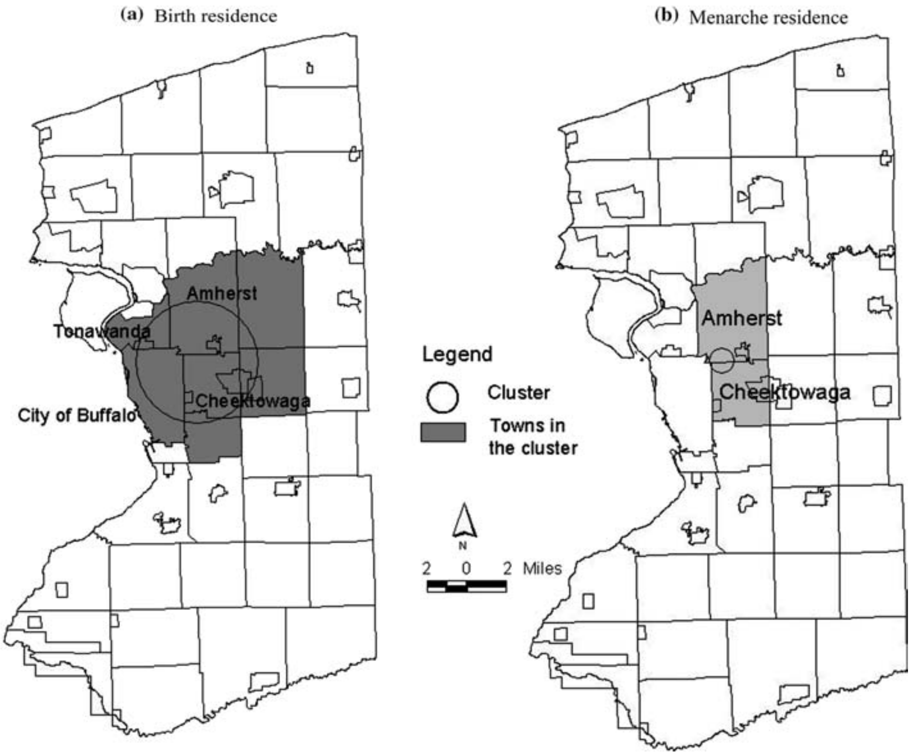

Fig. 4. Geographic clustering of residence at birth and menarche: pre-menopausal breast cancer, WEB Study, 1996–2001.

clustering of birth residence for pre-menopausal cases when compared to controls. We found a circular cluster of birth residence for breast cancer cases with a 5.7-mile radius in the area including part of the city of Buffalo, and the towns of Amherst, Cheektowaga, and Tonawanda (shaded areas). There are 100 observed breast cancer cases inside the cluster, while 76 breast cancer cases are expected. The cluster was significant at <0.01 with 999 Monte-Carlo simulations.

Further, we examined breast cancer risk associated with residence in the cluster at the time of birth. When we compared other breast cancer risk factors, such as age, education, and age at menarche, for the pre-menopausal breast cancer cases and controls whose birth residence was inside the cluster to those who lived outside of cluster, we did not find significant differences between the two groups (data not shown). We observed an elevated breast cancer risk for pre-menopausal women living in the cluster at the time of birth. With subjects living outside the cluster as a reference group, the adjusted odds ratio was 2.65 (95% CI 1.75–4.0) after controlling for age, education, age at menarche, parity, history of benign breast disease, and family history of breast cancer.

We also identified clustering of menarche residence for pre-menopausal women and obtained similar results as for birth residence. We were able to identify a small clustering of menarche residences for pre-menopausal breast cancer cases. A small cluster in the center of those four towns was detected (Figure 4b). It is a cluster with 0.8 mile radius and is statistically significant at $p < 0.05$. The cluster contains nine observed and 3.1 expected breast cancer cases, yielding a relative risk (ratio of observed to expected breast cancer cases) of 2.9. A secondary cluster was also detected near the city of Buffalo. It has a three-mile radius and relative risk of 1.38 with 65 observed and 47 expected breast cancer cases, but it is not statistically significant ($p = 0.38$).

Discussion

To our knowledge, no other studies have examined clustering of residential locations associated with cancer during early life: studies have examined clustering of residential locations at the time of diagnosis or death [20]. Critical time periods, including birth, menarche, and women's first pregnancy, as important early life and reproductive events in women's life, may play a substantial role in the risk of breast cancer. Under the hypothesis that there may be sensitive time periods in women's lives that will carry greater risk for exposure, the essential question was whether cases were more clustered than the underlying population, as represented by the controls. We found that cases were more clustered than controls at the time of birth and menarche, and it was due to clustering of residence for pre-menopausal, but not for post-menopausal breast cancer. The evidence for clustering of residential locations at birth and menarche was stronger than evidence for clustering at the time of women's first birth or other time periods in adult life. Our findings suggest that there may be identifiable etiological processes linking exposure and breast cancer risk, especially for pre-menopausal women, and that early exposures may be of particular importance.

This study provided a unique opportunity to examine clustering of breast cancer cases and controls at various points during early life. The facts that the study area had a relatively stable population and about 40% of study participants were lifetime residents, made the results more reliable. The evidence that residence in early life was important in the geographical clustering of breast cancer cases may be of particular importance for understanding environmental determinants of breast cancer. These findings suggest the importance of early or lifetime exposure in relation to disease risk in adult life, and also the potential role of the effects of migration on exposures and disease risk. Although migration can have a serious effect on the detection of geographical differences in disease risk, it has not been adequately addressed in previous clustering analyses [21]. Further investigations are required to prove any relationship between geographic clustering of residence and breast cancer risk, and the effects of residential changes on exposures should be considered in these studies.

Our finding of clustering was restricted to pre-menopausal breast cancer. We stratified on menopausal status because of evidence that there were differences in risk factors for pre- and post-menopausal women [22]. The mechanism of the observed difference is not clear. It could be that early life exposures impact pre-menopausal more than post-menopausal disease because of greater temporal proximity. There is some evidence, though not consistent, that other early exposures may differ by menopausal status. For example, there are data suggesting that birthweight may be more associated with pre- than with post-menopausal breast cancer [9, 23].

The results should be interpreted cautiously due to the fact that there may be some artifacts of the analysis. First, it is important to note that spatial point patterns are complex to summarize in a single way [24]. For example, the use of cumulative scales in the application of the *k*-function method may influence the outcome [25]. In particular, clustering is more likely to be detected on a larger geographic scale, and it tends to show continuous patterns over several neighboring scales due to the fact that the geographical scales are cumulative. Further refinement of methods to summarize spatial point patterns may provide more reliable results, as well as more accurate estimates of disease risk.

Second, this study is limited to current residents in the study area because we focused on the residential environment of Erie and Niagara counties; participants residing outside of these two counties at the time of each early life event were not included. The existence of missing residential information and potential selection bias due to non-participation may influence the results. As noted, we found no difference in participation by residence for cases compared to controls. Further we would expect that our findings on the clustering of early-life residence would be less subject to potential geographic selection bias than would current residence. We found a greater degree of clustering for residence at early life than for current residential location.

Further, the fact that residence at birth and menarche were often the same made it difficult to differentiate associations for the two time periods. For 22% of cases and 35% of controls, the menarche residence was the same as their birth residence. While the observed tendencies may be related to environmental exposures, it is also possible that clustering of residence at the time of birth or menarche may be due to clustering of other socioeconomic or demographic factors. Evaluation of the contribution of socioeconomic status to clustering of residences at birth and menarche is of special interest. There may be other factors associated with residence not measured in this study. The findings are still of interest for further study in order to understand what those exposures might be. We are now investigating the relation between spatio-temporal clustering of residences and exposures to environmental compounds, such as PAHs and benzene, to provide epidemiologic evidence of this finding.

Since the publication of John Snow's [26] well-known cholera map for the city of London in the 19th century, the relationship between the environment and disease has been one of the major research themes in medical geography. Geographic perspectives are of great use in describing geographical patterns of diseases, generating hypotheses on disease etiology, monitoring high risk areas of disease incidence, and suggesting possible causal factors of particular disease [27, 28]. Our study demonstrated that these GIS-based clustering analyses provide effective ways to explore spatial–temporal patterns of clustering. The findings show consistent results; the cluster identified by spatial analyses remained significant when traditional epidemiologic methods were used, and it was not explained by potential confounders. A recent study comparing 'traditional' epidemiological methods, GIS, and point pattern analysis for use in the spatially referenced public health data concluded that results complement, rather than contradict or duplicate each other [29].

In summary, this analysis of breast cancer clustering in space provides evidence of geographic clustering of pre-menopausal, but not post-menopausal, breast cancer cases at the time of birth and menarche, suggesting a possible influence of exogenous risk factors on breast cancer at these time points. While it is not clear from these data what caused this spatial clustering, it is provocative in providing evidence of the importance of this early period in breast carcinogenesis. Further investigations on genetic susceptibility may be of relevance to identify different effects on pre- and post-menopausal breast cancer. It will also be meaningful to see whether there is temporal clustering of early-life residences as well as spatial clustering. This type of study also needs to be replicated in other settings.

Acknowledgements

This work was supported in part by NIH Grants 1R01 ES09816-01, 1R21 CA87138-01, and U.S. Army Medical Research Grants DAMD17-03-1-0475, DAMD17-00-1-0417.

References

1. Sturgeon SR, Schairer CG, McAdams M, Brinton LA, Hoover RN (1995) Geographic variation in mortality from breast cancer among white women in the United States. *J Natl Cancer Inst* **87**: 1846–1853.
2. Lacey JV, Jr., Devesa SS, Brinton LA (2002) Recent trends in breast cancer incidence and mortality. *Environ Mol Mutagen* **39**: 82–88.
3. Wolff MS, Collman GW, Barrett C, Huff J (1996) Breast cancer and environmental risk factors: epidemiologic and experimental findings. *Annu Rev Pharmacol Toxicol* **36**: 573–596.
4. Laden F, Hunter DJ (1998) Environmental risk factors and female breast cancer. *Annu Rev Public Health* **19**: 101–123.
5. Kuh D, Ben-Shlomo Y (1997) A Life Course Approach to Chronic Disease Epidemiology: Tracing the Origins of Ill-health from Early to Later Life. Oxford: Oxford University Press.
6. Barker DJP (1992) *Fetal and Infant Origins of Adult Disease*. London: BMJ Publishing.
7. Colditz GA, Frazier AL (1995) Models of breast cancer show that risk is set by events of early life: prevention efforts must shift focus. *Cancer Epidemiol Biomarkers Prev* **4**: 567–571.
8. Trichopoulos D (1990) Hypothesis: does breast cancer originate *in-utero*? *Lancet* **335**: 939–940.
9. Potischman N, Troisi R (1999) *In-utero* and early life exposures in relation to risk of breast cancer. *Cancer Causes Control* **10**: 561–573.
10. Lumey LH (1998) Prenatal oestrogens and breast cancer. *Paediatr Perinat Epidemiol* **12**: 361–365.
11. Elliott P, Wakefield JC, Best NG, Briggs DJ (2000) *Spatial Epidemiology*: *Methods and Applications*. Oxford: Oxford University Press.
12. GDT (2001) GDT Dynamp/2000 Version 11.1, Geographic Data Technology, INC. USA.
13. Ripley BD (1981) *Spatial Statistics*. New York: Wiley.
14. Ripley BD (1977) Modelling spatial patterns. *J R Stat Soc Ser* B **39**: 172–212.
15. Rowlingson B, Diggle PJ (1993) SPLANCS: spatial point pattern analysis code in S-plus. *Comput Geosci* **19**: 627–655.
16. Diggle PJ, Chetwynd AD (1991) Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics* **47**: 1155–1163.
17. Kulldorff M, Nagarwalla N (1995) Spatial disease clusters: detection and inference. *Stat Med* **14**: 799–810.
18. Kulldorff M (1997) A spatial scan statistic. *Commun Stat-Theor M* **26**:1481–1496.
19. Bailey TC, Gatrell AC (1995) *Interactive Spatial Data Analysis*. Harlow: Longman.
20. Timander LM, McLafferty S (1998) Breast cancer in West Islip, NY: a spatial clustering analysis with covariates. *Soc Sci Med* **46**: 1623–1636.
21. Rogerson PA, Han D (2002) The effects of migration on the detection of geographic differences in disease risk. *Soc Sci Med* **55**: 1817–1828.
22. McPherson K, Steel CM, Dixon JM (2000) ABC of breast diseases: breast cancer epidemiology, risk factors, and genetics. *Br Med J* **321**: 624–628.
23. Stavola B, Hardy R, Kuh D, Sllva I, Wadsworth M, Swerdlow A (2000) Birthweight, childhood growth and risk of breast cancer in a British cohort. *Br J Cancer* **83**: 964–968.
24. Cressie N (1993) *Statistics for Spatial Data*. New York: Wiley.
25. Reader S (2000) Using survival analysis to study spatial point patterns in geographical epidemiology. *Soc Sci Med* **50**: 985–1000.
26. Snow J (1855) On the mode of communication of cholera. In: Snow on cholera being a reprint of two papers by John Snow, M.D. 1936. New York: The Commonwealth Fund.
27. Jones K, Moon G (1987) *Health, Disease, and Society*. London: Routledge.
28. Meade MS, Florin JW, Gesler WM (1988) *Medical Geography*. New York: Guilford.
29. Dunn C, Kingham S, Rowlingson B et al. (2001) Analysing spatially referenced public health data: a comparison of three methodological approaches. *Health Place* **7**: 1–12.